

学位論文 (要約)

Structural and functional analysis of innate immune receptors  
for double-stranded DNA and cyclic dinucleotide

(二本鎖 DNA・環状ジヌクレオチドに対する

自然免疫受容体の構造機能解析)

平成 29 年 12 月博士 (理学) 申請

東京大学大学院理学系研究科

生物科学専攻

大村 洋記

## Abstract

In the innate immune system, pattern recognition receptors (PRRs) specifically recognize ligands derived from bacteria or viruses, to trigger the responsible downstream pathways. DEAD box protein 41 (DDX41) is an intracellular PRR that triggers the downstream pathway involving the adapter STING, the kinase TBK1, and the transcription factor IRF3, to activate the type I interferon response. DDX41 is unique in that it recognizes two different ligands; *i.e.*, double-stranded DNA (dsDNA) and cyclic dinucleotides (CDN), *via* its DEAD domain. However, the structural basis for the ligand recognition by the DDX41 DEAD domain has remained elusive. In chapter 1, the author report two crystal structures of the DDX41 DEAD domain in apo forms, at 1.5 and 2.2 Å resolutions. A comparison of the two crystal structures revealed the flexibility in the ATP binding site, suggesting its formation upon ATP binding. Structure-guided functional analyses *in vitro* demonstrated the overlapped binding surface for dsDNA and CDN, which is distinct from the ATP-binding site. The author propose that the structural rearrangement of the ATP binding site is crucial for the release of ADP, enabling the fast turnover of DDX41 for the dsDNA/CDN-induced STING activation pathway. In chapter 2, the author determined the crystal structure of autoimmune disease-associated protein. The structure suggests the mechanism of causing diseases.

## Contents

<b>General Introduction</b> .....	<b>2</b>
<b>Chapter 1. Structural and Functional Analysis of DDX41: a bispecific immune receptor for double-stranded DNA and cyclic dinucleotide.</b> .....	<b>5</b>
<b>1.1 Introduction</b> .....	<b>5</b>
<b>1.2 Materials and methods</b> .....	<b>8</b>
Protein preparation.....	8
Crystallization .....	11
Data collection, structure determination and refinement .....	12
Pull-down assay .....	12
Kinase assay .....	13
<b>1.3 Results</b> .....	<b>15</b>
Expression and purification of DDX41 .....	15
The DDX41 DEAD domain recognizes both dsDNA and CDN .....	16
DDX41 recognizes dsDNA and CDN at the same binding site.....	19
Crystallization and structure determination of DDX41 .....	21
Overall structures of the DDX41 DEAD domain .....	24
Structural rearrangement of the ATP-binding site of the DDX41 DEAD domain .....	25
Putative dsDNA and CDN binding sites of DDX41 .....	29
<b>1.4 Discussion</b> .....	<b>33</b>
<b>1.5 Future works</b> .....	<b>39</b>
<b>Chapter 2.</b> .....	<b>40</b>
<b>General Discussion</b> .....	<b>41</b>
<b>References</b> .....	<b>43</b>
<b>Acknowledgements</b> .....	<b>48</b>

## General Introduction

The innate immune system functions as the first-line of defense against invading bacteria and viruses. In this system, proteins called pattern recognition receptors (PRRs) recognize pathogen-associated molecular patterns (PAMPs) as ligands, and finally induce the production of interferons and inflammatory cytokines, or activate the acquired immunity<sup>1</sup>. To date, five different classes of PRR families have been reported (Table 1)<sup>1</sup>; transmembrane PRRs including Toll-like receptors (TLRs)<sup>1,2</sup> and C-type lectin receptors (CLRs)<sup>3</sup>, as well as cytoplasmic PRRs including the human PYHIN family member absent in melanoma 2-like receptors (ALRs)<sup>4</sup>, the Retinoic acid-inducible gene (RIG)-I-like receptors (RLRs)<sup>5</sup> and nucleotide-binding and oligomerization domain (NOD) -like receptors (NLRs)<sup>6</sup>. Furthermore, recent studies also reported PRRs which are not classified into these five classes, such as cGAS<sup>7</sup>. Almost all PRRs share the common feature to specifically recognize only one type of ligand (Table 1). The ligand specificity is critical to distinguish self and non-self, namely to prevent to attack the host itself and to selectively attack bacteria and viruses. To reveal how PRRs retain its ligand specificity, structural and functional analyses have been performed<sup>8</sup>.

One of the PAMPs, double-stranded DNA (dsDNA), is the effective target of PRRs because of its wide conservation in all pathogenic bacteria and in dsDNA viruses. Extracellular or phagosomal bacteria leak or secrete dsDNA and/or cyclic dinucleotide (CDN) via secretion systems during its infection<sup>9,10</sup>. DsDNA viruses directly inject its genomic dsDNA into the host cells. RNA viruses trigger the reverse transcript of its RNA and release it to the host cells. Viral infection also triggers



mitochondrial DNA (mtDNA) stress, resulting in the release of mtDNA<sup>11,12</sup>. These pathogenic DNA or mtDNA is recognized by cytosolic DNA sensor(s), and triggers the production of type I interferon as well as the autophagy pathway, to induce the bacterial clearance. In this dsDNA dependent innate immune pathway, the ER-resident membrane protein STING is required to trigger the interferon production<sup>13,14</sup>. STING itself has low affinity for dsDNA<sup>15</sup>, suggesting that the direct dsDNA sensor(s) functions upstream of STING. Previous research identified the putative dsDNA sensors<sup>16</sup>; AIM2<sup>17-22</sup>, IFI16<sup>23-25</sup>, LRRFIP1<sup>26</sup>, DHX36<sup>27</sup>, DHX9<sup>27,28</sup>, Ku70<sup>29</sup>, DNA-PKcs<sup>30</sup>, MRE11<sup>31</sup>, Rad50<sup>32</sup>, DDX41<sup>33</sup> and cGAS<sup>7,34-37</sup>. Recent studies revealed the detailed mechanism how cGAS recognizes dsDNA and transduces the signal to downstream STING<sup>37-41</sup>. However, the mechanism how other putative dsDNA sensors transduce signals remains elusive.

In the chapter 1, the author focus on the DEAD box protein 41 (DDX41). DDX41 is the PRR which is unique in that it recognize two different types of ligands, dsDNA and CDN. To elucidate the ligand recognition mechanism, the author determined the crystal structure of DDX41. Furthermore, *in vitro* functional analysis based on the structure suggests the putative ligand recognition sites, and gave the insight into the unique ligand-bispecificity of DDX41. In the chapter 2, 刊行予定の未公開データを含むため, 博士論文の全文を公表できない場合のガイドラインに従い, 非公開とします.

**Table 1. PRRs and their ligands**

<b>PRRs</b>	<b>ligand</b>
<b>TLR</b>	
TLR1	Triacyl lipoprotein
TLR2	Lipoprotein
TLR3	dsRNA
TLR4	LPS
TLR5	Flagellin
TLR6	Diacyl lipoprotein
TLR7	ssRNA
TLR9	CpG-DNA
TLR11	Profilin-like molecule
<b>CLR</b>	
Dectin-1	$\beta$ -Glucan
Dectin-2	$\beta$ -Glucan
MINCLE	SAP130
<b>ALR</b>	
AIM2	dsDNA
IFI16	dsDNA
<b>RLR</b>	
RIG-I	dsRNA
MDA5	dsRNA
<b>NLR</b>	
NOD1	iE-DAP
NOD2	MDP

# **Chapter 1. Structural and Functional Analysis of DDX41: a bispecific immune receptor for double-stranded DNA and cyclic dinucleotide.**

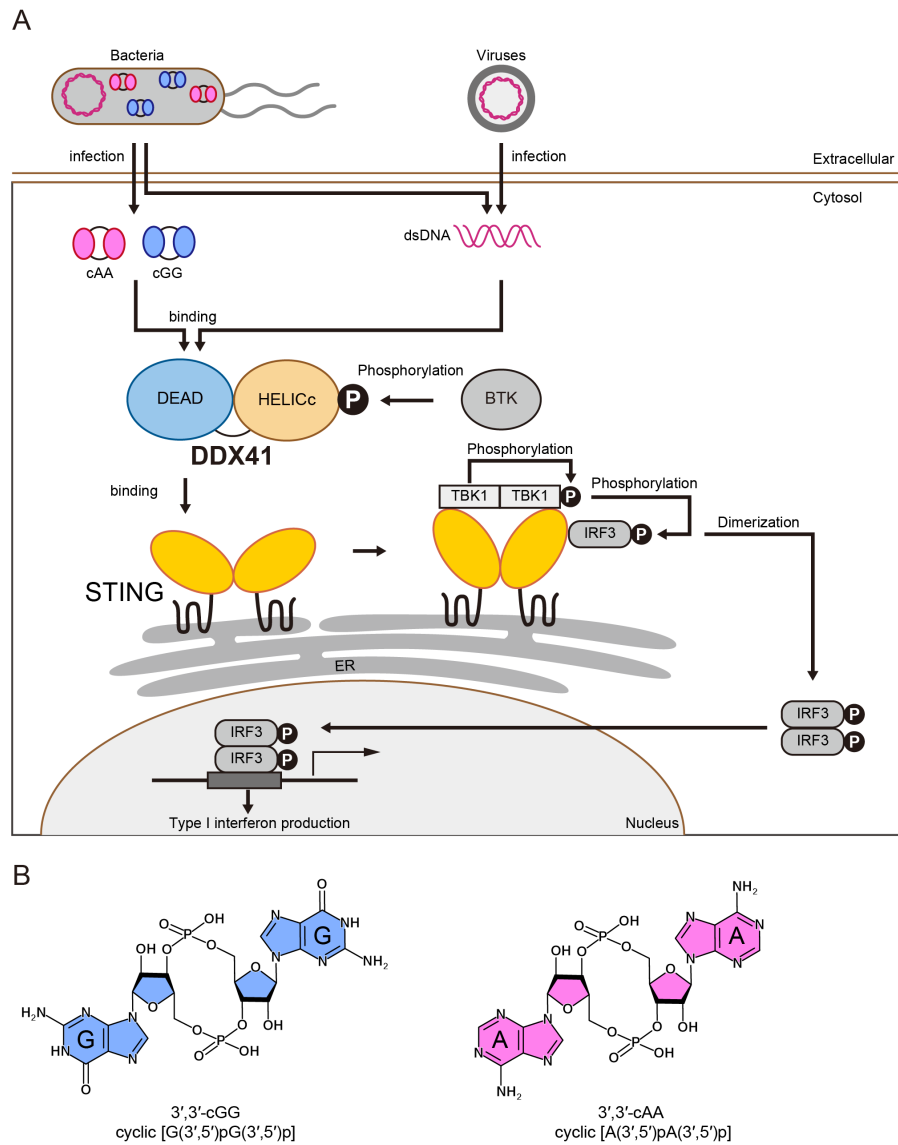
## **1.1 Introduction**

Almost all of the PRRs share the common feature of specifically recognizing a single type of PAMP. In contrast, DEAD box protein 41 (DDX41), a member of the helicase superfamily 2 (SF2), is unique in that it can recognize two different types of ligands, double-stranded DNA (dsDNA) and CDN<sup>33,42</sup>. DDX41 consists of two RecA-like domains (*i.e.*, DEAD and HELICc domains), which are conserved among DEAD box proteins. These domains have several conserved motifs, such as the Q motif and motif I, which are both essential for ATP binding by the DEAD box proteins<sup>43</sup>. A previous study showed that DDX41 containing the DEAD and HELICc domains has the ATPase activity<sup>44</sup>. The crystal structure of the DDX41 HELICc domain revealed its structural similarity to the HELICc domains of other DEAD box proteins<sup>45</sup>.

In the signaling pathway of the DDX41-related innate immune responses (Figure 1-1A), previous reports demonstrated that its HELICc domain is phosphorylated by Bruton's tyrosine kinase (BTK)<sup>46</sup>, and then it recognizes dsDNA and/or cyclicdinucleotide (CDN), including cyclic di-GMP (cGG) and cyclic di-AMP (cAA) (Figure 1-1B), via its DEAD domain<sup>33,42</sup>. After the ligand recognition, DDX41 binds to STING (stimulator of interferon genes) located on the endoplasmic reticulum

membrane<sup>33,42</sup>. STING subsequently promotes the phosphorylation of TBK1 (TANK-Binding Kinase 1) and IRF3 (interferon regulatory transcription factor 3), which finally induce the production of type I interferon<sup>13,47</sup>.

To understand the molecular mechanism of this signaling pathway, the structural basis for the mechanism of dsDNA and CDN recognition by the DDX41 DEAD domain is required. Although previous studies showed that dsDNA disturbs the DDX41–CDN interaction<sup>42</sup>, the binding site of the ligands and the mechanism by which DDX41 recognizes them have remained elusive. In this work, the author determined two different crystal structures of the DDX41 DEAD domain in apo forms. Comparisons between these two structures and the AMPPNP-bound structure of another DEAD box protein revealed the flexibility in the ATP binding site, suggesting its formation upon ATP binding. In addition, structure-guided functional analyses revealed the residues critical for dsDNA and CDN recognition, and suggested the putative dsDNA/CDN-binding surface.



**Figure 1-1. The signaling pathway of DDX41.** (A) Overview of the DDX41-STING-dependent signaling pathway. DDX41 is activated through phosphorylation on its HELICc domain by BTK. Phosphorylated DDX41 then recognizes dsDNA and/or CDN and activates the endoplasmic reticulum (ER)-resident receptor, STING. STING recruits TBK1 and IRF3, and induces the phosphorylation of them. Phosphorylated IRF3 is dimerized and translocates into the nucleus, to induce the production of type I interferon. (B) Structural formulae of the cyclic dinucleotides; cyclic [G(3',5')pG(3',5')p] (cGG, left panel), cyclic [A(3',5')pA(3',5')p] (cAA, right panel).

## 1.2 Materials and methods

### Protein preparation

The gene encoding *Homo sapiens* DDX41 (residues 1–622, FL-DDX41) was inserted into the modified pET28a vector (Novagen), and then FL-DDX41 or its truncated variants (DEAD, DEAD-HELICc, DEAD-HELICc-ZF or HELICc) were subcloned into the pE-SUMOpro Kan vector (LifeSensors). Point mutations of DEAD (residues 169–402) were introduced by QuikChange Site-Directed Mutagenesis (Agilent Technologies). All DDX41 constructs were expressed as N-terminal His<sub>6</sub>-SUMO-tag fused proteins in *E. coli* Rosetta2 (DE3) (Novagen) cells. The cells were grown at 37°C in LB medium to an OD<sub>600</sub> of 0.8. After the induction of protein expression by 0.4 mM isopropyl-β-D-thiogalactopyranoside, the cells were further cultured at 20°C for 18 h and then harvested by centrifugation.

All constructs of DDX41, except for HELICc, were purified in a similar manner, as follows. The harvested cells were resuspended in buffer A (50 mM HEPES-NaOH, pH 7.0, 300 mM NaCl, 20 mM imidazole, 10% glycerol, 2 mM MgCl<sub>2</sub>, 5 mM 2-mercaptoethanol) containing benzonase nuclease (Novagen), and were lysed by sonication. The lysates were centrifuged at 40,000g for 30 min, and then the supernatants were loaded onto Ni-NTA Superflow resin (QIAGEN) packed in an Econo-Column (Bio-Rad). The DDX41-bound resin was washed with buffer A, and then the proteins were eluted with buffer B (50 mM HEPES-NaOH, pH 7.0, 300 mM NaCl, 500 mM imidazole, 10% glycerol, 2 mM

MgCl<sub>2</sub>, 5 mM 2-mercaptoethanol). The eluted proteins were dialyzed with His-tag fused Ulp1 protease (prepared in our lab) at 4°C for 20 h, to remove the His<sub>6</sub>-SUMO-tag of DDX41. The proteins were again loaded onto Ni-NTA Superflow resin (QIAGEN) packed in an Econo-Column (Bio-Rad). The flow-through and wash fractions were collected, and then were loaded onto a HiTrap Heparin HP Column (GE Healthcare), equilibrated with buffer C (50 mM HEPES-NaOH, pH 7.0, 100 mM NaCl, 10% glycerol, 2 mM MgCl<sub>2</sub>, 1 mM DTT). The proteins were eluted using a linear gradient of 100–500 mM NaCl. The eluted proteins were diluted, and then loaded onto Resource S column (GE Healthcare) equilibrated with buffer C. The proteins were eluted using a linear gradient of 100–500 mM NaCl, and were concentrated using Amicon Ultra Centrifugal Filter Units (MWCO 10 kDa for DEAD WT and mutants, and 30 kDa for DEAD-HELICc, DEAD-HELICc-ZF and FL-DDX41) (Millipore). The proteins were then loaded onto a Superdex 200 Increase 10/300 size exclusion column (GE Healthcare) equilibrated with buffer D (50 mM HEPES-NaOH, pH 7.5, 150 mM NaCl, 10% glycerol, 2 mM MgCl<sub>2</sub>, 1 mM DTT). The proteins were concentrated with Amicon Ultra Centrifugal Filter Units (Millipore), flash frozen in liquid nitrogen, and stored at -80°C.

HELICc was purified in a similar manner to the other constructs, until the 2nd Ni-NTA column step. After purification on the Ni-NTA resin, the collected HELICc was loaded onto a HiTrap Heparin HP column (GE Healthcare), equilibrated with buffer C. The unbound fraction was collected and loaded onto a Resource Q column (GE Healthcare) equilibrated with buffer C. The unbound fraction was collected and concentrated using Amicon Ultra Centrifugal Filter Units (MWCO 10 kDa)

(Millipore). The protein was then loaded onto a Superdex 200 Increase 10/300 size exclusion column (GE Healthcare) equilibrated with buffer D. The protein was concentrated with Amicon Ultra Centrifugal Filter Units (MWCO 10 kDa) (Millipore), flash frozen in liquid nitrogen, and stored at -80°C.

The gene encoding *Mus musculus* BTK (residues 214–659) was cloned into the pFastBac HTb vector (Invitrogen) to create the baculovirus. Sf9 cells were infected with the baculovirus to express the N-terminal His<sub>6</sub>-tag- and Tobacco Etch Virus (TEV) protease cleavage site-fused BTK. The cells were harvested by centrifugation, 48 h after infection. The harvested cells were resuspended in buffer E (50 mM Tris-HCl, pH 8.0, 300 mM NaCl, 20 mM imidazole, 5 mM 2-mercaptoethanol), containing complete protease inhibitor (Roche), and lysed by sonication. The lysate was centrifuged at 40,000g for 30 min, and then the supernatant was further centrifuged at 138,000g for 1 h. The supernatant was loaded onto Ni-NTA Superflow resin (QIAGEN) packed in an Econo-Column (Bio-Rad). The BTK-bound resin was washed with buffer E, and then the protein was eluted with buffer F (50 mM Tris-HCl, pH 8.0, 300 mM NaCl, 500 mM imidazole, 5 mM 2-mercaptoethanol). The eluted protein was dialyzed with His-tag fused TEV protease (prepared in our lab) at 4°C for 20 h, to remove the His<sub>6</sub>-tag of BTK. The protein was loaded again onto Ni-NTA Superflow resin (QIAGEN) packed in an Econo-Column (Bio-Rad). The flow-through and wash fractions were collected and loaded onto a Resource Q column (GE Healthcare), equilibrated with buffer G (20 mM Tris-HCl, pH 8.5, 20 mM NaCl, 1 mM DTT). The protein was eluted as two separate peaks, using a linear gradient of 20–500 mM NaCl. These two



peaks correspond to the monomer and dimer of BTK, as previously described<sup>48</sup>. The fractions from the two peaks were collected separately, and were concentrated using Amicon Ultra Centrifugal Filter Units (MWCO 10 kDa) (Millipore). The protein was loaded onto a Superdex 200 Increase 10/300 size exclusion column (GE Healthcare) equilibrated with buffer H (20 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1 mM DTT). The protein was concentrated with Amicon Ultra Centrifugal Filter Units (MWCO 10 kDa) (Millipore), flash frozen in liquid nitrogen, and stored at -80°C.

## **Crystallization**

Two DDX41 constructs, DDX41<sub>169-402</sub> (residues 169–402) and DDX41<sub>169-399</sub> (residues 169–399), were used for crystallization. Before crystallization, DDX41<sub>169-402</sub> and DDX41<sub>169-399</sub> were diluted to 15 mg ml<sup>-1</sup> and 7.5 mg ml<sup>-1</sup>, respectively. DDX41<sub>169-402</sub> was then incubated with 5 mM cyclic di-GMP for 1 h at 4°C. After the incubation, DDX41<sub>169-402</sub> was highly precipitated. The precipitate was removed with an Ultrafree-MC Centrifugal Hydrophilic Filter Unit (Millipore). Crystallization trials were performed by the sitting drop vapor diffusion method at 4°C, using a Mosquito crystallization robot (TTP Labtech). Crystals of DDX41<sub>169-402</sub> and DDX41<sub>169-399</sub> were obtained with reservoir solution A (2.1 M DL-malic acid, pH 7.0) and reservoir solution B (0.18 M tri-ammonium citrate, 20 % (w/v) polyethylene glycol 3350), respectively. The reservoir solutions supplemented with 5 % and 25% glycerol were used as cryoprotectants for DDX41<sub>169-402</sub> and DDX41<sub>169-399</sub>, respectively.

## Data collection, structure determination and refinement

Diffraction data sets of DDX41<sub>169-402</sub> and DDX41<sub>169-399</sub> were collected at the Swiss Lightsource PXII and SPring-8 BL32XU, respectively. The data sets of DDX41<sub>169-402</sub> were processed with the programs XDS<sup>49</sup>. The data sets of DDX41<sub>169-399</sub> were processed with MOSFLM<sup>50</sup>, FECKLESS and AIMLESS<sup>51</sup>. The DDX41<sub>169-399</sub> crystals exhibited non-merohedral twinning. Two overlapping lattices were identified by the multiple-lattice indexing algorithm in MOSFLM and integrated separately, combined by FECKLESS, and scaled with AIMLESS. The phases of DDX41<sub>169-402</sub> were determined with the program Phaser<sup>52</sup>. The search model for DDX41<sub>169-402</sub> was created, based on the structure of the DDX5 DEAD domain (PDB ID: 3FE2)<sup>45</sup> and the amino-acid sequence of DDX41, using the program Molrep<sup>53</sup>. The phases of DDX41<sub>169-399</sub> were determined by the program Phaser, using the structure of DDX41<sub>169-402</sub> as the search model. The initial models were built using the program PHENIX<sup>54</sup>. For the model building and further refinement, the programs COOT<sup>55</sup>, PHENIX and Refmac5<sup>56</sup> were used.

## Pull-down assay

The sequence of the sense strand of ISD is as follows: 5'-TACAGATCTACTAGTGATCTATGACTGATCTGTACATGATCTACA-3'. To make the double-stranded bio-ISD, the 5' biotinylated sense strand of ISD (Eurofin) was mixed with its complementary

unlabeled DNA (Eurofin) and annealed. For negative controls and competition assays, unlabeled ATP, ADP, AMPPNP (SIGMA) and cGG (C057-01, BioLog) were used. Bio-ISD, bio-cGG (2'-O- (6-[biotinyl]aminohexylcarbamoyl)-cyclic diguanosine monophosphate, B098-001, BioLog), bio-ATP (N<sup>6</sup>-(6-amino)hexyl-adenosine-5'-triphosphate-biotin, NU-805-BIO, Jena Bioscience) and bio-AMPPNP (2'/3'-O-(2-aminoethyl-carbamoyl)-adenosine-5'-[( $\beta,\gamma$ )-imido] triphosphate-biotin, NU-810-BIO, Jena Bioscience), each at a final concentration of 1  $\mu$ M, were incubated with Dynabeads M-280 Streptavidin (Thermo Fisher Scientific) for 30 min at 4°C in binding buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 10% glycerol, 0.5 mM EDTA, 1 mM 2-mercaptoethanol, 0.5% Nonidet P-40). The beads were washed with binding buffer three times, and then were incubated with the proteins for 1 h at 4°C in binding buffer. The beads were washed with binding buffer three times. The bound proteins were eluted by boiling for 5 min at 95°C, using SDS-PAGE sample buffer.

## **Kinase assay**

Purified FL-DDX41 was incubated with the BTK dimer at a 20:1 molar ratio for 16 h at 4 °C, in the presence of 10 mM MgCl<sub>2</sub> and 1 mM ATP. To isolate the phosphorylated FL-DDX41, the protein was loaded onto a HiTrap Heparin HP column (GE Healthcare), equilibrated with buffer C (used in the purification of DDX41). The protein was eluted with a linear gradient of 100–1,000 mM NaCl. The protein was concentrated with Amicon Ultra Centrifugal Filter Units (MWCO 30 kDa) (Millipore), flash frozen in liquid nitrogen, and stored at -80°C. The phosphorylation of DDX41 and the self-

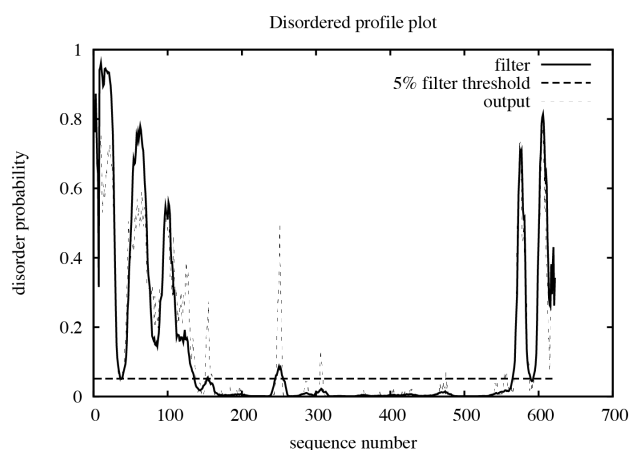
phosphorylation of BTK were detected by western blotting, using the anti-phosphotyrosine (4G10) antibody (05-321, Upstate Biotechnology), or by SuperSep Phos-tag (Wako). The phosphorylation of FL-DDX41 was detected, although those of DEAD-HELICc and DEAD-HELICc-ZF could not be detected.

## 1.3 Results

### Expression and purification of DDX41

To design the constructs of DDX41 suitable for crystallization and functional analysis, the author performed the secondary-structure prediction and disordered prediction, using programs Psipred<sup>57</sup> and Disopred<sup>58</sup>, respectively. The results showed that N-terminus (residues 1-160) and C-terminus (residues 600-622) of DDX41 is disordered (Figure 1-2A). To stably express DDX41, the author designed 5 constructs and tried the large-scale expression (Figure 1-2B). To improve the solubility, all constructs were expressed as N-terminal His<sub>6</sub>-SUMO tag fused proteins. The proteins were purified using Ni-NTA affinity column, heparin affinity column, ion exchange column and size exclusion column. The elute peak of size exclusion column exhibited tailing.

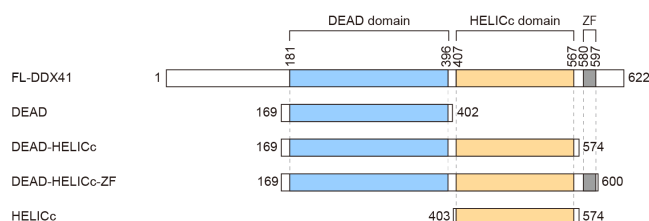
A



## Figure 1-2. Characteristics of DDX41.

(A) The disorder profile plots of FL-DDX41. The disordered prediction was performed using programs Disopred58. (B) FL-DDX41 and its truncated variants. DDX41 contains the DEAD domain (sky blue), the HELICc domain (orange) and the Zinc finger (ZF, gray). The color codes representing the domains and regions are used similarly in the other figures, unless otherwise stated.

B



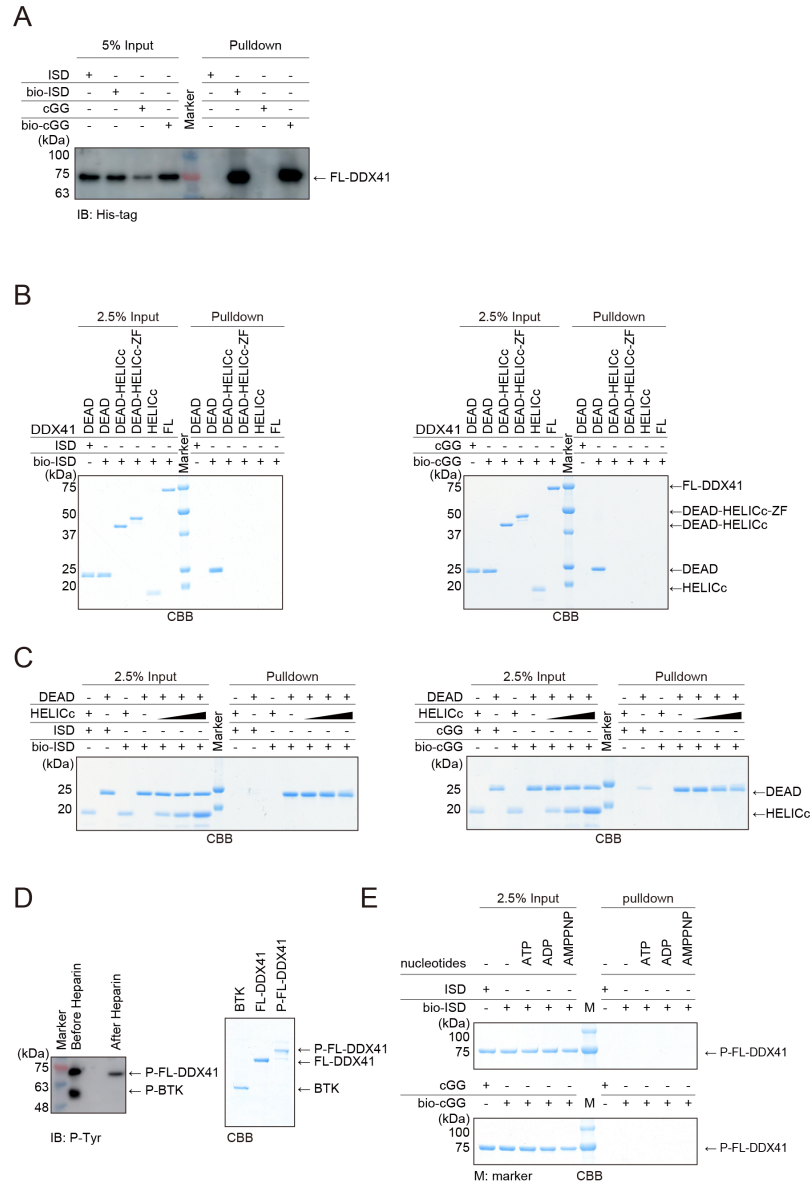
## The DDX41 DEAD domain recognizes both dsDNA and CDN

To understand the ligand recognition mechanism of DDX41, the author tested the dsDNA- and CDN-binding abilities of DDX41 and its variants. Previous studies reported that full-length DDX41 (FL-DDX41) and its truncated variants containing the DEAD domain (DEAD and DEAD-HELICc; Figure 1-2B) can bind to biotinylated dsDNA and CDN<sup>33,42</sup>. Therefore, the author examined the dsDNA- and CDN-binding abilities of FL-DDX41 expressed in HEK293T cells by pull-down assays, using 5'-biotinylated immune stimulatory DNA (bio-ISD) or biotinylated cyclic di-GMP (bio-cGG) (Figure 1-3A). The results showed that FL-DDX41 expressed in HEK293T cells bound both bio-ISD and bio-cGG. Next, the author examined dsDNA- and CDN-binding abilities of purified FL-DDX41

and its truncated variants (DEAD, DEAD-HELICc, DEAD-HELICc-ZF, and HELICc; Figure 1-2B) by pull-down assays, using 5'-biotinylated immune stimulatory DNA (bio-ISD) or biotinylated cyclic di-GMP (bio-cGG). The results showed that DEAD can bind to both ISD and cGG, whereas neither DEAD-HELICc, DEAD-HELICc-ZF nor FL-DDX41 bound to either ISD or cGG (Figure 1-3B). These results indicated that the DEAD domain itself can bind to both dsDNA and CDN, and its binding activity is inhibited by the presence of the HELICc domain *in vitro*.

To further assess the possibility that the HELICc domain inhibits the dsDNA- and CDN-binding activities of the DEAD domain, the author tested the effect of the isolated HELICc domain on the dsDNA- and CDN-binding activities of DEAD. The results showed that increasing amounts of HELICc did not affect the dsDNA- and CDN-binding activities of DEAD (Figure 1-3C). These results suggested that the linker between the DEAD and HELICc domains is important for the inhibitory activity. This linkage anchors the HELICc domain to raise its local concentration around the DEAD domain, where it may mask the dsDNA- and CDN-binding surface on the DEAD domain.

Previous studies reported that the BTK-dependent phosphorylation of the HELICc domain is crucial for the activation and ligand recognition of DDX41<sup>46</sup>. To clarify the inhibition mechanism of the dsDNA- and CDN-binding activities by the HELICc domain, the author examined the ligand-binding abilities of phosphorylated FL-DDX41 by the pull-down assay. However, the results showed that phosphorylated FL-DDX41 bound neither dsDNA nor CDN, regardless of the presence and absence of ATP, ADP and AMPPNP (Figure 1-3D and 1-3E).



**Figure 1-3. The DDX41 DEAD domain binds to dsDNA and CDN.** (A) Pull-down assays of FL-DDX41 expressed in HEK293T cells. The pull-downed DDX41 was detected via C-terminal His6 tag by Western blotting, using anti-His-tag antibody (PM032, Medical & Biological Laboratories). (B) Pull-down assays of DDX41, using bio-ISD (left panel) and bio-cGG (right panel). In all of the pull-down assays, the final concentration of bio-ISD, bio-cGG, bio-ATP or bio-AMPPNP is 1  $\mu$ M, unless otherwise stated. Gels were run under the same experimental conditions and are shown as cropped gels. (C) Pull-down assay of DEAD, using bio-ISD (left panel) and bio-cGG (right panel) in the presence of HELICc. (D) Phosphorylation assay of DDX41 by BTK. FL-DDX41 was mixed with BTK and incubated in the presence of ATP and MgCl<sub>2</sub>. The phosphorylation of tyrosine was detected by an immunoblot (left panel, “Before Heparin”). Phosphorylated FL-DDX41 (P-FL-DDX41) was isolated using a HiTrap Heparin HP column (left panel, “After Heparin”). The isolated P-FL-DDX41 was detected by a mobility shift assay, using a SuperSep Phos-tag gel (right panel). (E) Pull-down assay of phosphorylated FL-DDX41, using bio-ISD (upper panel) and bio-cGG (left panel).

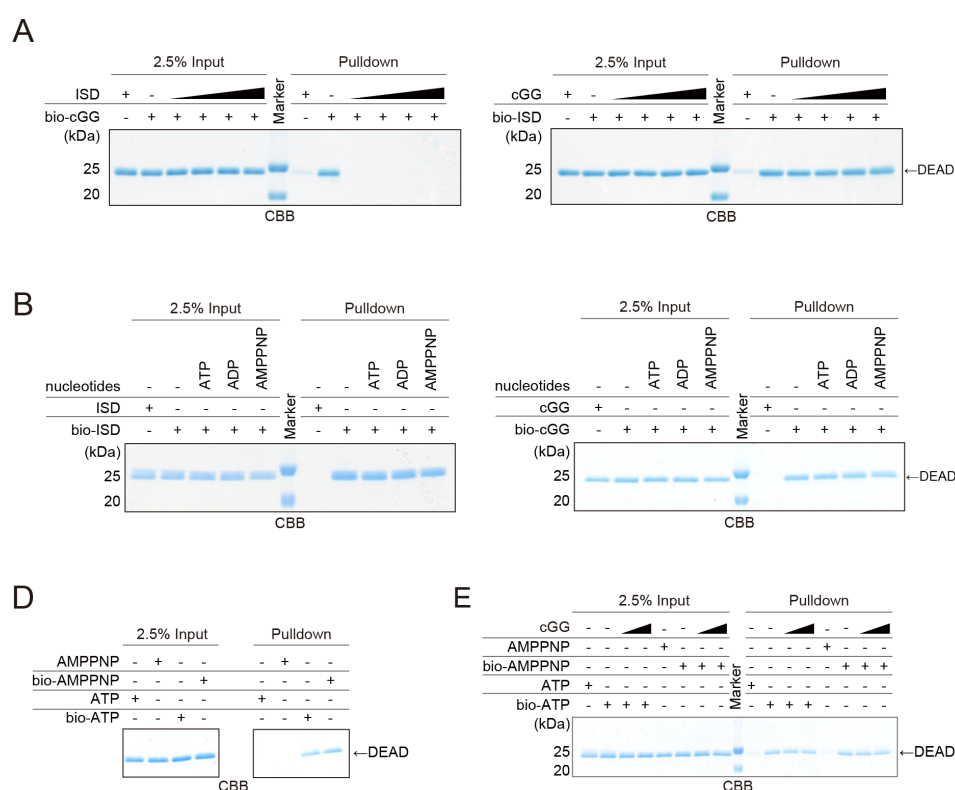


## **DDX41 recognizes dsDNA and CDN at the same binding site**

To further explore the dsDNA, CDN and ATP recognition mechanism by DDX41, the author examined whether dsDNA binding to the DEAD domain is affected by CDN, and *vice versa*, using the pull-down assay. The results revealed that the interaction between DEAD and bio-cGG was strongly impaired by the presence of increasing amounts of unlabeled ISD (Figure 1-4A, left panel), indicating that dsDNA binding competes with CDN binding to the DDX41 DEAD domain. In contrast, the interaction between DEAD and bio-ISD was not affected by the presence of increasing amounts of unlabeled cGG (Figure 1-4A, right panel). Thus, these results suggested that both dsDNA and CDN are recognized at the same ligand-binding site of the DEAD domain, which has higher affinity for dsDNA than CDN.

The DEAD box proteins typically bind to ATP via the conserved ATP-binding site, which is mainly located in the DEAD domain<sup>43</sup>. Several crystal structures of other DEAD box proteins bound to ADP, AMP or AMPPNP have been reported. Thus, the author first examined the ATP-binding ability of the DEAD domain by the pull-down assay, using biotinylated ATP (bio-ATP) or biotinylated AMPPNP (bio-AMPPNP), and determined that DEAD binds to both bio-ATP and bio-AMPPNP (Figure 1-4D). Next, the author examined whether ATP, ADP, and AMPPNP affect the binding of dsDNA and CDN to the DEAD domain, and *vice versa*. The author found that the presence of ATP, ADP or AMPPNP had no effect on both the bio-cGG and bio-ISD binding (Figure 1-4B). Similarly, the presence of unlabeled cGG had no effect on both the bio-ATP and bio-AMPPNP binding (Figure

1-4E). Collectively, these results suggested that the recognition sites for dsDNA and CDN are different from that for ATP, even though ATP and CDN share common structural features, including the base, ribose, and phosphate moieties.

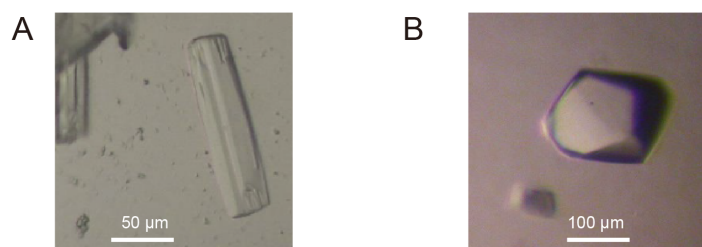


**Figure 1-4. dsDNA and CDN bind to the DEAD domain at the same sites.** (A) Pull-down assays of DEAD, using bio-cGG in the presence of unlabeled ISD (left panel), and bio-ISD in the presence of unlabeled cGG (right panel). The final concentrations of unlabeled ISD or cGG are 0, 0.5, 1, 2 or 4  $\mu$ M. (B) Pull-down assays of DEAD, using bio-ISD (left panel) or bio-cGG (right panel), in the presence of 1 mM unlabeled ATP, ADP or AMPPNP. Gels were run under the same experimental conditions and are shown as cropped gels. (C) Pull-down assay of DEAD, using bio-ATP or bio-AMPPNP. (D) Pull-down assay of DEAD, using bio-ATP or bio-AMPPNP, in the presence of 0, 1 or 4  $\mu$ M unlabeled cGG.

## Crystallization and structure determination of DDX41

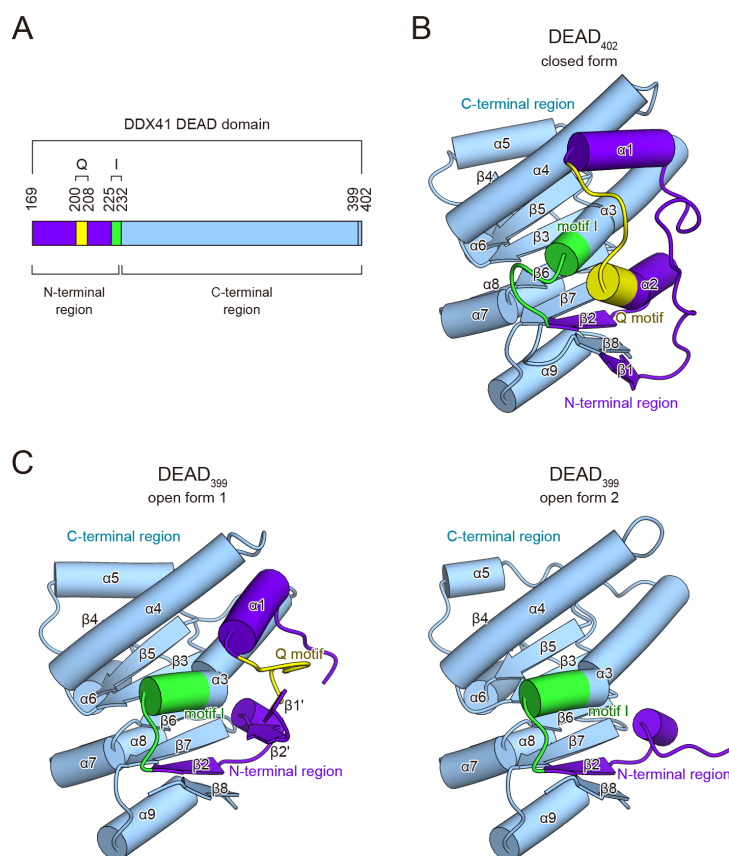
To obtain insights into the dsDNA- and CDN-recognition mechanism of DDX41, the author performed the crystallization screening, using the purified DDX41, in the presence or absence of AMPPNP and cGG. The author obtained the crystal, using the construct of DDX41 containing residues 169-399 (DDX41<sub>169-399</sub>) in the absence of any ligands with reservoir solution containing 0.18 M triammonium citrate, 20 % (w/v) polyethylene glycol 3350. To obtain larger crystal, the author optimized the crystallization condition, and finally obtained the crystal of over 100  $\mu\text{m}$  (Figure 1-5A). To obtain the diffraction data, the author performed the X-ray diffraction experiments using the crystal, at SPring8 BL32XU. The crystal exhibited the non-merohedral twin, therefore unable to determine the phase. To solve the twinning by changing the packing of the crystal, the author designed the N- and C-terminus deletion and elongation constructs. To obtain crystals, the author expressed, purified and crystallized these constructs in the same way as DDX41<sub>169-399</sub>. The author obtained the crystal of DDX41<sub>169-402</sub>, in the different crystallization condition from that of DDX41<sub>169-399</sub>, containing 2.1 M DL-malic acid, pH 7.0 (Figure 1-5B). To obtain the diffraction data, the author performed the X-ray diffraction experiments using the crystal, at Swiss Lightsource PXII. To determine the initial model, the author processed the data sets, and performed molecular replacement. Furthermore, to obtain the another crystal form, the author performed the detwinning of DDX41<sub>169-399</sub>, using the structure of DDX41<sub>169-402</sub> as a search model. After refinement of the initial models, the author determined the crystal structure of DDX41<sub>169-402</sub> and DDX41<sub>169-399</sub> at 1.5 and 2.2 Å resolutions, respectively<sup>59</sup> (Figure

1-6A–1-6C, Table 1-1).



**Figure 1-5. The crystals of DDX41.**

(A) The crystal of DDX41169-399. (B) The crystal of DDX41169-402.



**Figure 1-6. Overall structures of the DEAD domain.**

(A) The motif organization of the DEAD domain. The DEAD domain contains the N-terminal region (violet), the C-terminal region (sky blue), the Q motif (yellow) and motif I (green). The color codes representing the motifs and the N- or C-terminal regions are used similarly in the other figures, unless otherwise stated. The molecular graphics in the all figures were rendered using the program CueMol (<http://www.cuemol.org/>). (B) The overall structure of the closed form. (C) The overall structures of the open form 1 (left panel) and the open form 2 (right panel).

**Table 1-1. Data Collection and Refinement Statistics of DDX41**

	DDX41 DEAD (169-402)	DDX41 DEAD (169-399)
<b>Data collection</b>		
Beamline	SLS PXII	SPring-8 BL41XU
Wavelength (Å)	0.979	1.000
Space group	<i>P</i> 3 <sub>2</sub> 21	<i>I</i> 2
Cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	81.83, 81.83, 69.57	91.19 50.96 202.60
$\alpha$ , $\beta$ , $\gamma$ (°)	90, 90, 120	90 95.52 90
Resolution (Å)	40.92–1.50	49.41–2.20
<i>R</i> <sub>meas</sub>	0.065 (>1.00)	0.202 (0.867)
<i>I</i> / $\sigma$ <i>I</i>	18.40 (1.47)	9.3 (2.8)
Completeness (%)	99.9 (99.5)	76.5 (77.5)
Redundancy	10.34 (10.24)	4.4 (4.1)
<b>Refinement</b>		
Resolution (Å)	40.92–1.50	49.41–2.20
No. reflections	41,380	36,167
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub>	0.184/0.211	0.207/0.275
No. atoms		
Protein	1826	6473
Ligand	9	0
Solvent	114	333
<i>B</i> -factors		
Protein	29.66	24.49
Ligand	35.92	-
Solvent	31.44	21.67
R.m.s. deviations		
Bond lengths (Å)	0.023	0.013
Bond angles (°)	2.18	1.696
Ramachandran plot		
Favored (%)	97.41	96.69
Allowed (%)	2.59	3.19
Outlier (%)	0	0.12

\*The highest resolution shell is shown in parentheses.

## Overall structures of the DDX41 DEAD domain

The determined crystal structures of DDX41<sub>169-402</sub> and DDX41<sub>169-399</sub> exhibited the apo forms, although DDX41<sub>169-402</sub> was crystallized in the presence of cGG. Both structures adopt a RecA-like fold with one core  $\beta$ -sheet surrounded by  $\alpha$ -helices, and their C-terminal regions (233–399) have almost the same conformation, with an RMSD of 1.16 Å over 161 C $\alpha$  atoms (Figure 1-6B and 1-6C). Thus, when the author focus on the C-terminal region, the author discusses the DDX41<sub>169-402</sub> structure, since it was determined at higher resolution. In contrast, the conformations of the N-terminal region (169–232) are different in the crystal structures. In the crystal structure of DDX41<sub>169-402</sub>, helix  $\alpha$ 2 and the Q motif are located adjacent to helix  $\alpha$ 3 and motif I, and the N-terminal residues (169–173) form a  $\beta$  strand ( $\beta$ 1) and participate in the core  $\beta$ -sheet formation (Figure 1-3B). The author hereafter refers to this conformation of the N-terminal region as the “closed” form. In contrast to DDX41<sub>169-402</sub>, the crystal structure of DDX41<sub>169-399</sub> contains four molecules in the asymmetric unit, and their N-terminal regions adopt two different conformations (Figure 1-6C). In one conformation, the Q motif is located away from  $\alpha$ 3 and motif I. Helix  $\alpha$ 2 is rearranged to form two  $\beta$  strands ( $\beta$ 1' and  $\beta$ 2'), and residues 169–183 are disordered (Figure 1-6C, left panel). The author hereafter refers to this conformation of the N-terminal region as the “open form 1”. In the other conformation, helix  $\alpha$ 3 and motif I adopt a similar conformation to those in open form 1, while residues 169–209, including the entire Q motif, are disordered (Figure 1-6C, right panel). The author hereafter refers to this conformation of the N-terminal region as the “open form 2”.

## Structural rearrangement of the ATP-binding site of the DDX41 DEAD domain

Several crystal structures of DEAD box proteins in the ADP-, AMP- or AMPPNP-bound forms have been reported, and suggested that the important motifs for ATP recognition, including the Q motif and motif I, are structurally conserved<sup>45</sup>. Among the DEAD box proteins, the amino-acid sequence of DDX3 is most similar to that of DDX41 (the amino-acid sequence identity is 40% overall, 37% in the DEAD domain and 44% in the HELICc domain ). To understand the ATP recognition mechanism by the DDX41 DEAD domain, the author compared the present crystal structures with that of the DDX3 DEAD domain in the AMPPNP-bound form (PDB ID: 5E7M).

The closed form of the DDX41 DEAD domain superimposes well on that of DDX3, with an RMSD of 1.56 Å over 215 Cα atoms. Especially, this superimposition revealed that the structure around the ATP-binding site, including the Q motif and motif I, adopts a similar conformation to that of DDX3 (Figure 1-7A). This similarity between these proteins allowed us to create a docking model of the DDX41-ATP complex, based on the DDX3-AMPPNP complex structure (Figure 1-7B). In the crystal structure of the DDX3-AMPPNP complex, Tyr200 recognizes the adenine base of AMPPNP via a stacking interaction, while the corresponding residue is Ile (Ile201) in DDX41, suggesting that the adenine base is recognized by a hydrophobic interaction in DDX41 (Figure 1-7B). Similar base recognition by an Ile residue was also observed in the crystal structure of the DDX53-AMP complex, supporting our structural model (Figure 1-7A, PDB ID: 3IUY<sup>45</sup>). Furthermore, our structural model suggested that Gln208 in the Q motif of DDX41 recognizes the adenine moiety in a base-specific

manner, similar to Gln207 in DDX3 (Figure 1-7B, right panel). The Q208S mutant of DEAD exhibited decreased ATP-binding ability as compared to the wild-type protein (Figure 1-7C), suggesting the importance of Gln208 for the nucleotide binding. Furthermore, the model suggested that the main chain amide groups and Lys231 in motif I extensively recognize the  $\beta$ - and  $\gamma$ -phosphate groups of ATP, in a similar manner to those of motif I in DDX3 (Figure 1-7B, 1-8A and 1-8B).

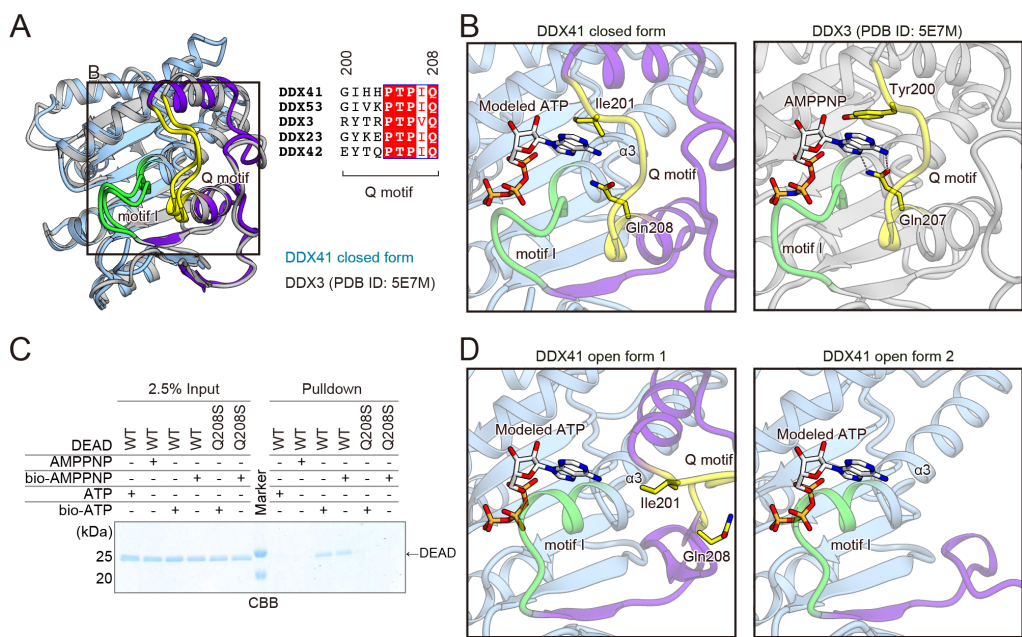
Next, the author created docking models of ATP-bound DDX41, based on the open form structures (Figure 1-7D). The models suggested that the structural rearrangement and disorder observed in the open forms 1 and 2 may affect the ATP binding by the DEAD domain. In these models, the Q motif, including Ile201 and Gln208, is distant from the modeled ATP in the open form 1, while it is disordered in the open form 2 (Figure 1-7D). Moreover, in both open forms 1 and 2, motif I forms part of the extended  $\alpha 3$  helix, in both open forms 1 and 2, which sterically clashes with the modeled ATP (Figures 1-7D, 1-8C and 1-8D).

Interestingly, a malic acid molecule from the crystallization solution is present at this putative binding site for the ATP phosphate group in the closed form structure (Figure 1-9). One of the two carboxylate groups of the malic acid interacts with the main chain amide groups of motif I, while the other carboxylate group forms a salt bridge with the Arg residues in the adjacent molecule in the crystalline lattice. This observation is consistent with the fact that the crystals of the closed form were obtained only under conditions containing malic acid.

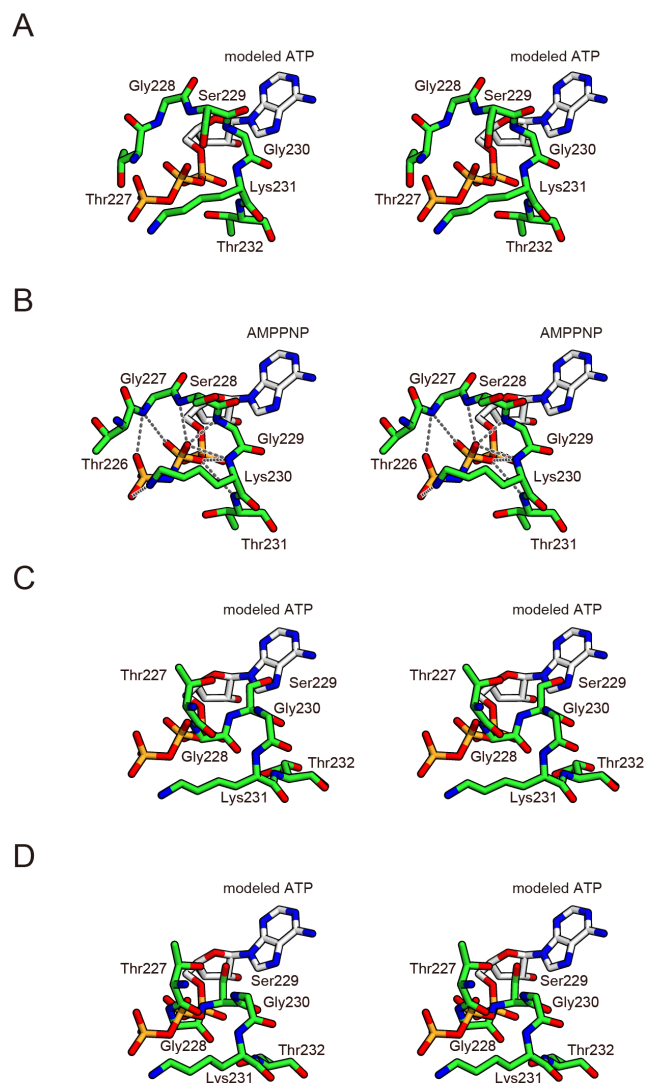
Consequently, the structure of the ATP-binding site is completely rearranged in these open



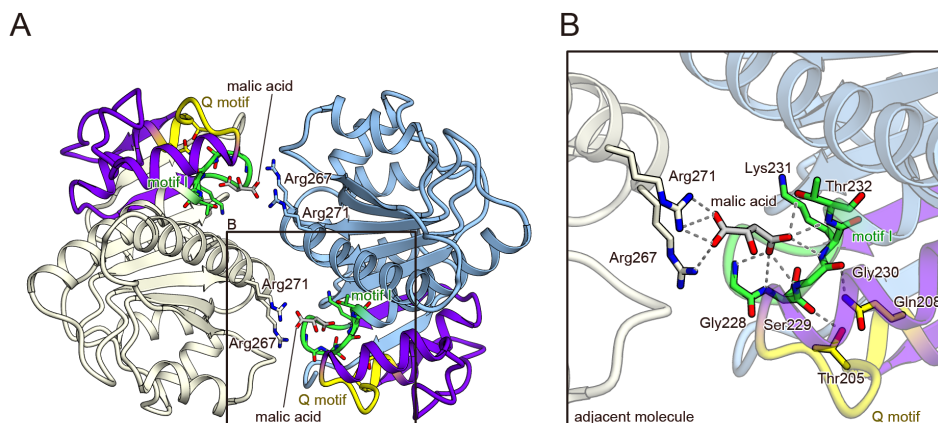
forms, and thus they do not seem to be able to bind ATP. Therefore, these results suggested that the closed form represents the ATP-bound state of DDX41, while both open forms 1 and 2 represent its nucleotide-free state.



**Figure 1-7. Three forms of DDX41 reveal the structural rearrangement in the ATP-binding site.** (A) Structure and sequence comparison between DDX41 and DDX3 (PDB ID: 5E7M). The Q motif and motif I of DDX3 are colored as in DDX41, whereas the other regions are colored silver. The amino-acid sequences of full-length DDX23 and DDX42 are the second and third most similar to that of DDX41. (B) Docking model of the DDX41 closed form bound to ATP (left panel), based on the DDX3-AMPPNP complex structure (PDB ID: 5E7M) (right panel). (C) Pull-down assays of the DEAD mutant, using bio-ATP and bio-AMPPNP. A cropped gel is shown. (D) Docking models of the DDX41 open forms bound to ATP, based on the DDX3-AMPPNP complex structure (PDB ID: 5E7M). Models of the open form 1 (left panel) and the open form 2 (right panel) are shown.



**Figure 1-8. Motif I of the DDX41 closed form interacts with modeled ATP in a similar manner to that of DDX3.** (A)(B)(C)(D) The ATP binding mode of motif I. The motif I structures shown (stereo views) are the closed form (A), DDX3 (PDB ID: 5E7M) (B), the open form 1 (C) and the open form 2 (D).

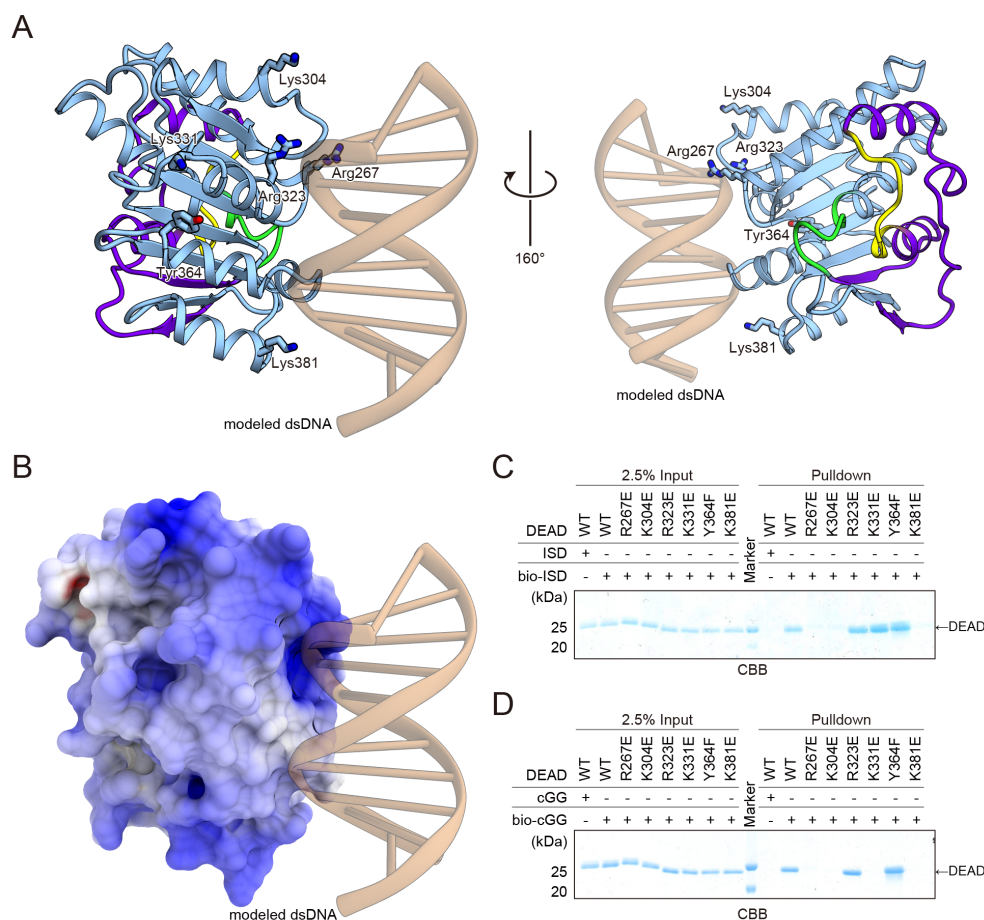


**Figure 1-9. Malic acid binds to motif I of DDX41 by mimicking the phosphate group of ATP. (A)** Interaction between the closed form of DDX41 and malic acid. The adjacent molecule of DDX41 is colored beige. **(B)** The details of the interaction between DDX41 and malic acid.

## Putative dsDNA and CDN binding sites of DDX41

To obtain structural insight into its dsDNA-binding mechanism, the author created a model structure of the dsDNA-bound DEAD domain of DDX41. Among the SF2 proteins with structures reported in complex with dsDNA, the RecA-like domain of Rad54<sup>60</sup> shares the highest structural similarity (23.9% sequence identity) with the DDX41 DEAD domain. The closed form of the DDX41 DEAD domain was superimposed on that of Rad54, with an RMSD of 2.36 Å over 157 Cα atoms. In addition, a previous study<sup>33</sup>, as well as our data (Figure 1-2B), indicated that the HELICc domain of DDX41 is not directly involved in the dsDNA binding. Thus, the author created the docking model of the DDX41 DEAD domain bound to dsDNA, based on the Rad54 structure (PDB ID: 1Z63) (Figure

1-10A and 1-10B). The model suggested that the dsDNA binding site is located around the surface of the C-terminal region, and involves Arg267, Lys304, Tyr364 and Lys381 (Figure 1-10A). To corroborate this putative dsDNA-binding site, the author examined the effects of mutations in this site on the dsDNA-binding ability, by the pull-down assay using bio-ISD. The results showed that the R267E, K304E and K381E mutations decrease the bio-ISD binding ability of DEAD (Figure 1-10C), suggesting the importance of Arg267, Lys304 and Lys381 for the dsDNA binding. Next, the author examined the CDN-binding ability of these mutants, using bio-cGG. The results demonstrated that the R267E, K304E and K381E mutations also decreased the CDN-binding ability of DEAD (Figure 1-10D), thus suggesting that these residues in the putative dsDNA-binding surface are involved in both dsDNA and CDN binding. In contrast, the K331E mutation exhibited decreased binding ability only for CDN and not dsDNA (Figure 1-10C and 1-10D), implying that this residue is exclusively involved in CDN binding. Thus, these results suggested that the dsDNA- and CDN-binding sites of DDX41 overlap with each other. This notion is further supported by a previous report that bio-cGG binding to FL-DDX41 is affected by unlabeled dsDNA<sup>42</sup>. We hereafter refer to this surface around the C-terminal region as the putative dsDNA/CDN-binding surface. Furthermore, Lys304, Lys331 and Lys381 on this dsDNA/CDN-binding surface are not conserved in other paralogues of DDX41, such as DDX3, DDX23, DDX42, and DDX53 (Figure 1-11A), while they are highly conserved in the orthologues of DDX41 from other species (Figure 1-11B). These observations are consistent with the fact that the dsDNA- and CDN-binding activities are specific features of DDX41.



**Figure 1-10. Putative dsDNA and CDN binding sites of DDX41.** (A) Docking model of the DDX41 DEAD domain bound to dsDNA. The dsDNA is colored brown. (B) Electrostatic surface potentials of the DDX41 closed form (contoured from  $-10 \text{ kT e}^{-1}$  [red] to  $+10 \text{ kT e}^{-1}$  [blue]). (C) Pull-down assays of the DEAD mutants and bio-ISD. (D) Pull-down assays of the DEAD mutants and bio-cGG. Gels were run under the same experimental conditions and are shown as cropped gels.



## 1.4 Discussion

DDX41 is a unique PRR that recognizes two different types of PAMPs: dsDNA and CDN<sup>33,42</sup>.

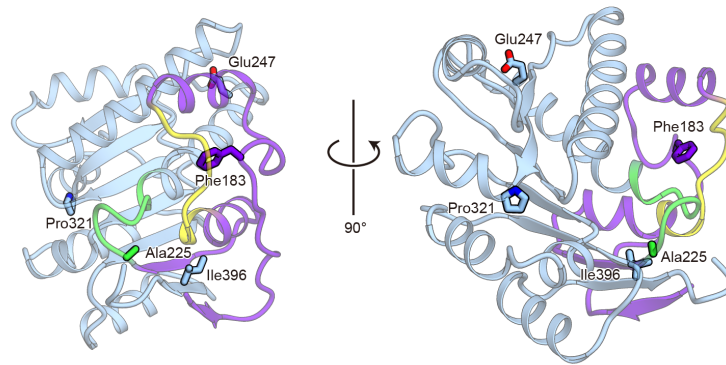
In this study, the author determined the crystal structures of the DDX41 DEAD domain in the closed and open forms. Previous studies reported that the mutations of residues in the DEAD domain of human DDX41, F183I, A225D, E247K, P321L and I396T, cause acute myeloid leukemia (AML) syndrome<sup>61</sup>. All of these residues form the inside core of the DEAD domain (Figure 1-12), suggesting that the mutations of these residues lead to the misfolding of DDX41, which causes AML. Based on the apo form structure, the author created the model structure of dsDNA-bound DDX41 and performed the mutational analyses. These analyses revealed the residues involved in the dsDNA and CDN binding, and suggested that the DEAD domain recognizes these different ligands at overlapping sites.

After the author reported the crystal structures of DDX41 in *Scientific Reports* journal, Jiang *et al* also reported the crystal structure of the DDX41 DEAD domain in the apo form, containing residues 149-408<sup>62</sup>. The overall structure of DDX41<sub>149-408</sub> has almost the same conformation as the DDX41<sub>169-402</sub> closed form with an RMSD of 0.52 Å over 230 Cα atoms, except the N-terminal additional helix “α0” (residues 149-169) (Figure 1-13).

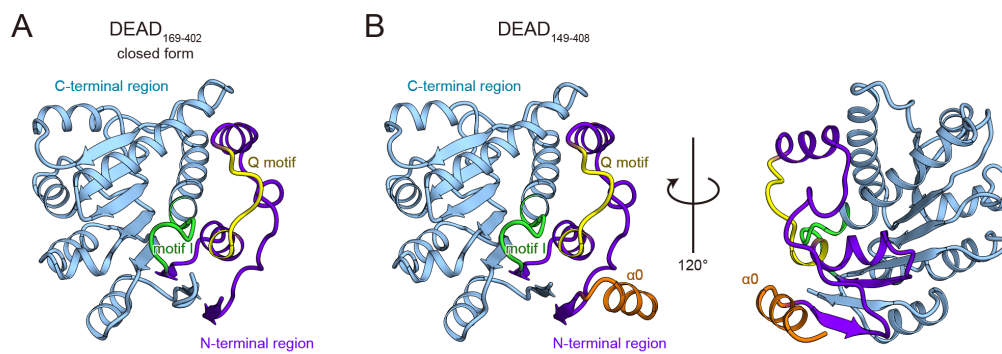
A structural comparison between the open and closed forms of the DDX41 DEAD domain revealed the structural rearrangement in the N-terminal region, which drastically changes the conformation of the ATP-binding site. One of the intriguing points of these structures is that the ATP-binding site is formed in the closed form without binding an adenine nucleotide. Instead, the

carboxylate group of malic acid contained in the crystallization solution is bound to the phosphate-binding site formed by motif I (Figure 1-9). Thus, in the present crystal structure, the DDX41 DEAD domain is trapped in the closed form by the carboxylate group of malic acid, which is bound by mimicking the phosphate group. This observation led to a putative mechanism of the structural change that occurs in the DDX41 DEAD domain upon ATP binding. The phosphate moiety of ATP binds to motif I, which induces the helix-to-loop transition of motif I (Figure 1-7B, 1-7D and 1-8). This structural transition of motif I enables the interaction between the Q motif and motif I, including the hydrogen bonds between Thr205, Gln208 in the Q motif and Ser229, Gly230 in motif I (Figure 1-9). These interactions fix the Q motif, including Gln208 and Ile201, in the closed-form conformation and thus create the adenine moiety binding pocket (Figure 1-7B). In the cellular context, this structural change may occur in the transition from the closed form in the ATP-bound state to the open form after the ATP hydrolysis and subsequent ADP release. What is the role of this structural change of DDX41 in the dsDNA and CDN sensing mechanism? RIG-I (DDX58), another SF2 protein that functions as a cytosolic PRR, recognizes single- and double-stranded RNA to trigger the downstream pathway for the innate immune responses<sup>1</sup>. Previous studies reported that ATP binding and hydrolysis and ADP release accelerate the dissociation of the bound RNA to facilitate the fast ligand recognition turnover of RIG-I<sup>63,64</sup>. Thus, it is possible that the structural transition between the open and closed forms of DDX41 also accelerates the binding and release of ligands to facilitate the fast turnover, as in the case of RIG-I.





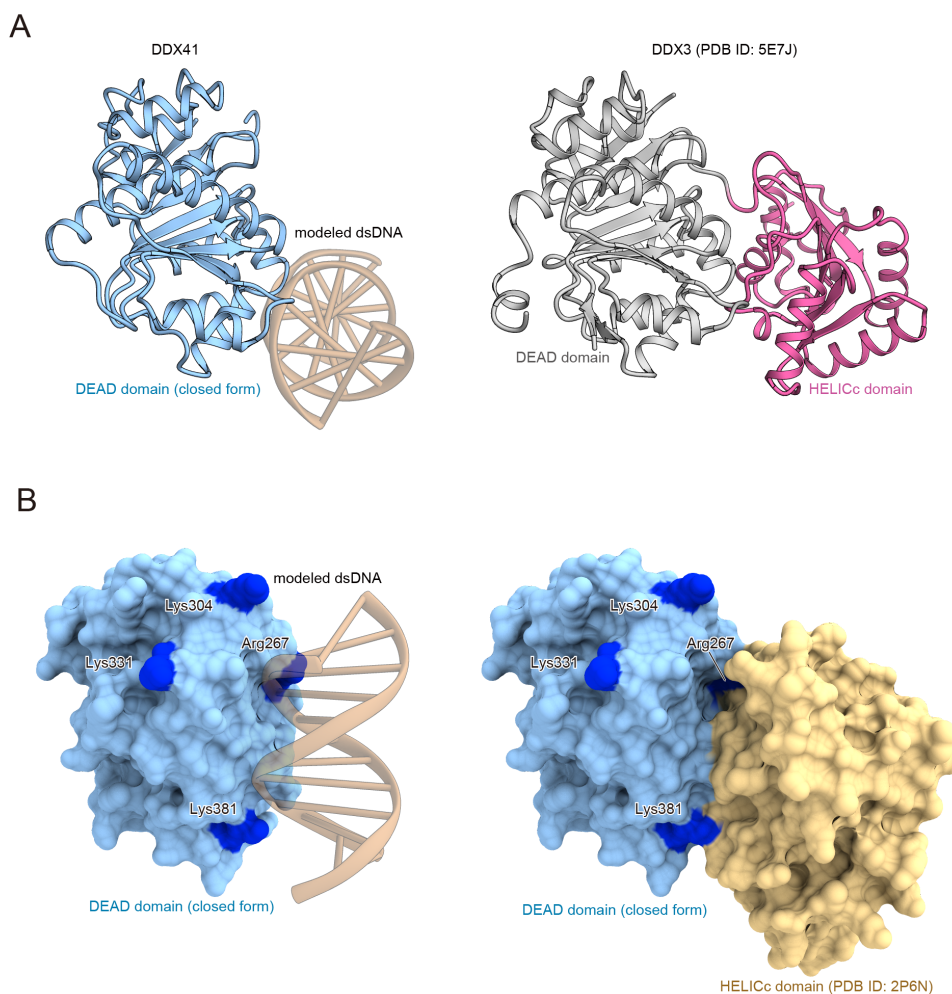
**Figure 1-12. Residues mutated in AML patients form the inside core of the DEAD domain.** The residues mutated in AML patients are mapped on the structure of the DEAD domain.



**Figure 1-13. Structural comparison between the DDX41169-402 closed form and DDX41149-408.** (A) The crystal structure of DDX41169-402, in the closed form. (B) The crystal structure of DDX41149-408 determined by Jiang et al., 2017. The additional helix  $\alpha 0$  is colored in orange.

The results of our pull-down assays using the purified proteins prepared by the *E. coli* expression system demonstrated that the DDX41 variants containing only the DEAD domain bind to dsDNA and CDN, while those containing both the DEAD and HELICc domains do not (Figure 1-2B). These results suggested that the HELICc domain inhibits the interaction between the DEAD domain and these ligands. Given the crystal structure of the DDX41 HELICc domain<sup>45</sup> and the length and location of the linker region between the DEAD and HELICc domains, the HELICc domain can adopt a conformation where it is located near the dsDNA/CDN-binding surface, to disturb its interactions with ligands. For example, in the DDX41 DEAD-HELICc model structure created based on the crystal structure of DDX3<sup>65</sup> (Figure 1-14A), the dsDNA/CDN-binding surface of the DEAD domain is completely covered by the HELICc domain (Figure 1-14B). Furthermore, our pull-down assays demonstrated that the isolated HELICc domain does not affect the ligand binding ability of the DEAD domain, suggesting that the interaction between the DEAD and HELICc domains is weak. The linker region between the DEAD and HELICc domains restricts the location of the HELICc domain and keeps it near the dsDNA/CDN-binding surface, which may perturb the ligand binding by the DEAD domain. In contrast, previous studies, using a crude extract prepared from cultured human cells, showed that FL-DDX41 also binds to both biotinylated dsDNA and CDN<sup>33,42</sup>. Furthermore, in a luciferase reporter assay in L929 cells, the HELICc-truncated variant reportedly exhibited higher *Irfn*-promoter induction activity upon dsDNA recognition than that of FL-DDX41<sup>33</sup>. Thus, these results suggest the possibility that an unidentified factor or post-translational modification disturbs the

interaction between the HELICc and DEAD domains, which allows ligand access to the dsDNA/CDN-binding surface accessible for ligands. Another PRR, TLR (toll-like receptor) 4, requires a MD-2 protein as a co-factor to recognize its ligand, LPS (lipopolysaccharide)<sup>66,67</sup>. In this study, the author examined the possibility that the phosphorylation by BTK is involved in the activation of FL-DDX41, but found that it had no effect on the ligand binding (Figure 1-3D and 1-3E). Overall, in conjunction with the previous reports, our present structural and functional analyses strongly suggest the existence of a regulation mechanism of DDX41, although further studies will be required to identify the regulatory factor(s).



**Figure 1-14. Models of HELICc domain-mediated inhibition of dsDNA and CDN binding by the DEAD domain.** (A) Docking model of DDX41 bound to dsDNA (left panel), and the structure of DDX3 DEAD-HELICc (PDB ID: 5E7J) (right panel). The DEAD domain of DDX41 is colored sky blue. The DEAD domain and the HELICc domain of DDX3 are colored silver and magenta, respectively. The two structures are in the same orientation. (B) Docking model of DDX41 bound to dsDNA (left panel) and model of DDX41 DEAD-HELICc (right panel), in the molecular surface representation. The latter model is based on the structure of DDX3 (PDB ID: 5E7J). Residues in the putative dsDNA/CDN-binding surface are colored blue. The DEAD domain and the HELICc domain (PDB ID: 2P6N) are colored sky blue and orange, respectively. The two models are in the same orientation.

## 1.5 Future works

Although the author suggested the putative ligands binding sites, the detail of ligand recognition mechanism still remains unclear. To elucidate the mechanism, the crystal structures of DDX41-CDN complex and DDX41-dsDNA complex are required. The author tries to obtain the crystals of complexes by co-crystallization and soaking methods, and to determine the structures.

The results of the binding assays suggest the possibility that an unidentified factor or post-translational modification disturbs the interaction between the HELICc and DEAD domains. The author tries to identify the factor or modification. Previous research revealed the binding partners of DDX41 in the non-phosphorylated form by mass-spectrometry<sup>68</sup>. The author tries to identify the binding partner(s) of phosphorylated DDX41 by mass-spectrometry. Comparison of the binding molecules to the non-phosphorylated- and phosphorylated-DDX41 reveals the molecule(s) which specifically interact with phosphorylated-DDX41, which may be the activator of DDX41. To prove the molecule as the activator, the author performs the binding assay of DDX41 to ligands in the presence of the molecules. The author elucidate the activation mechanism of DDX41 by the activator.

## Chapter 2.

刊行予定の未公開データを含むため、博士論文の全文を公表できない場合のガイドラインに従い、非公開とします。

## General Discussion

In the chapter 1, the author performed the structural and functional analysis of bispecific PRR, DDX41. Almost all PRRs share the feature to recognize only one type of ligands. The correspondence of one PRR for one type of ligand is the simplest strategy to distinguish host and invading pathogens. In contrast, DDX41 recognize two different types of ligands. The structural and functional analysis suggest that DDX41 recognizes both dsDNA and CDN in the same ligand recognition sites in the competitive manner, and the affinity for dsDNA is greater than that for CDN. However, a lot of questions remain elusive in the DDX41-STING pathway. How DDX41 transduce the signal to STING? Is the signal transduction mechanism the same in both dsDNA- and CDN-pathway? What is the cofactor of DDX41 to activate it? To deal with these questions, further structural and functional analyses are required.

Although many DNA sensors working upstream of STING have been reported<sup>16</sup>, the function of these receptors are controversial, except cGAS, which is well-known and is considered as the main PRR for dsDNA in STING-dependent pathway<sup>69</sup>. In this research, the author revealed the DNA- and CDN-binding capacity of DDX41 *in vitro*, but the detailed function of DDX41 *in vivo* remains elusive, especially in terms of the relationship to cGAS. One possibility, the specificity for DNA length and sequence, or the cells in which they work, may differ between cGAS and DDX41. Another possibility is that DDX41 functions as the intermediate adaptor, which carries and localizes dsDNA to cGAS, or present CDN to STING, rather than the direct sensor for dsDNA and CDN. Further functional analysis

*in vivo* is required in the context of collaboration or habitat-segregation between cGAS and DDX41.

In the chapter 2, 刊行予定の未公開データを含むため, 博士論文の全文を公表できない  
場合のガイドラインに従い, 非公開とします.



## References

1. Takeuchi, O. & Akira, S. Pattern recognition receptors and inflammation. *Cell* **140**, 805–820 (2010).
2. Akira, S., Uematsu, S. & Takeuchi, O. Pathogen recognition and innate immunity. *Cell* **124**, 783–801 (2006).
3. Geijtenbeek, T. B. H. & Gringhuis, S. I. Signalling through C-type lectin receptors: Shaping immune responses. *Nat. Rev. Immunol.* **9**, 465–479 (2009).
4. Ratsimandresy, R. A., Dorfleutner, A. & Stehlik, C. An update on PYRIN domain-containing pattern recognition receptors: From immunity to pathology. *Front Immunol.* **4**, (2013).
5. Yoneyama, M. & Fujita, T. Structural Mechanism of RNA Recognition by the RIG-I-like Receptors. *Immunity* **29**, 178–181 (2008).
6. Zhong, Y., Kinio, A. & Saleh, M. Functions of NOD-Like Receptors in Human Diseases. *Front. Immunol.* **4**, (2013).
7. Sun, L., Wu, J., Du, F., Chen, X. & Chen, Z. J. Cyclic GMP-AMP synthase is a cytosolic DNA sensor that activates the type I interferon pathway. *Science (80-. ).* **339**, 786–91 (2013).
8. Yin, Q., Fu, T.-M., Li, J. & Wu, H. Structural Biology of Innate Immunity. *Annu Rev Immunol* **33**, 393–416 (2015).
9. Stetson, D. B. & Medzhitov, R. Recognition of Cytosolic DNA Activates an IRF3-Dependent Innate Immune Response. *Immunity.* **24**, 93–103 (2006).
10. Monroe, K. M., Mcwhirter, S. M. & Vance, R. E. Induction of type I interferons by bacteria. *Cell Microbiol.* **12**, 881–890 (2010).
11. Fang, C., Wei, X. & Wei, Y. Mitochondrial DNA in the regulation of innate immune responses. *Protein Cell* **7**, 11–6 (2016).
12. West, A. P. *et al.* Mitochondrial DNA stress primes the antiviral innate immune response. *Nature* **520**, 553–7 (2015).
13. Ishikawa, H. & Barber, G. N. STING is an endoplasmic reticulum adaptor that facilitates innate immune signalling. *Nature* **455**, 674–678 (2008).
14. Ishikawa, H., Ma, Z. & Barber, G. N. STING regulates intracellular DNA-mediated , type I interferon-dependent innate immunity. *Nature* **461**, 788–792 (2009).
15. Abe, T. *et al.* STING Recognition of Cytoplasmic DNA Instigates Cellular Defense. *Mol. Cell* **50**, 5–15 (2013).
16. Dempsey, A. & Bowie, A. G. Innate Immune Recognition of DNA : a recent history. *Virology* 146–152 (2015). doi:10.1016/j.virol.2015.03.013.Innate
17. Hornung, V. *et al.* AIM2 recognizes cytosolic dsDNA and forms a caspase-1-activating inflammasome with ASC. *Nature* **458**, 514–518 (2009).
18. Roberts, T. L. *et al.* HIN-200 Proteins Regulate Caspase Activation in Response to Foreign Cytoplasmic DNA. *Science (80-. ).* **323**, 1057–1060 (2009).
19. Bürckstümmer, T. *et al.* An orthogonal proteomic-genomic screen identifies AIM2 as a cytoplasmic DNA sensor for the inflammasome. *Nat. Immunol.* **10**, 266–272 (2009).

20. Fernandes-Alnemri, T., Yu, J. W., Datta, P., Wu, J. & Alnemri, E. S. AIM2 activates the inflammasome and cell death in response to cytoplasmic DNA. *Nature* **458**, 509–513 (2009).
21. Fernandes-Alnemri, T. *et al.* The AIM2 inflammasome is critical for innate immunity to *Francisella tularensis*. *Nat. Immunol.* **11**, 385–393 (2010).
22. Rathinam, V. A. K. *et al.* The AIM2 inflammasome is essential for host defense against cytosolic bacteria and DNA viruses. *Nat. Immunol.* **11**, 395–402 (2010).
23. Unterholzner, L. *et al.* IFI16 is an innate immune sensor for intracellular DNA. *Nat. Immunol.* **11**, 997–1004 (2010).
24. Horan, K. A. *et al.* Proteasomal Degradation of Herpes Simplex Virus Capsids in Macrophages Releases DNA to the Cytosol for Recognition by DNA Sensors. *J Immunol.* **190**, 2311–2319 (2013).
25. Jin, T. *et al.* Structures of the HIN Domain: DNA Complexes Reveal Ligand Binding and Activation Mechanisms of the AIM2 Inflammasome and IFI16 Receptor. *Immunity* **36**, 561–571 (2012).
26. Yang, P. *et al.* The cytosolic nucleic acid sensor LRRFIP1 mediates the production of type I interferon via a beta-catenin-dependent pathway. *Nat. Immunol.* **11**, 487–494 (2010).
27. Kim, T. *et al.* Aspartate-glutamate-alanine-histidine box motif (DEAH)/RNA helicase A helicases sense microbial DNA in human plasmacytoid dendritic cells. *Proc. Natl. Acad. Sci.* **107**, 15181–15186 (2010).
28. Zhang, Z., Yuan, B., Lu, N., Facchinetti, V. & Liu, Y. DHX9 pairs with IPS-1 to sense double-stranded RNA in myeloid dendritic cells. *J Immunol* **187**, 4501–4508 (2011).
29. Zhang, X. *et al.* Ku70 is a novel cytosolic DNA sensor that induces type III rather than type I IFN. *J Immunol* **186**, 4541–4545 (2011).
30. Ferguson, B. J., Mansur, D. S., Peters, N. E., Ren, H. & Smith, G. L. DNA-PK is a DNA sensor for IRF-3-dependent innate immunity. *Elife* **2012**, e00047 (2012).
31. Kondo, T. *et al.* DNA damage sensor MRE11 recognizes cytosolic double-stranded DNA and induces type I interferon by regulating STING trafficking. *Proc. Natl. Acad. Sci.* **110**, 2969–2974 (2013).
32. Roth, S. *et al.* Rad50-CARD9 interactions link cytosolic DNA sensing to IL-1 $\beta$  production. *Nat. Immunol.* **15**, 538–545 (2014).
33. Zhang, Z. *et al.* The helicase DDX41 senses intracellular DNA mediated by the adaptor STING in dendritic cells. *Nat Immunol* **12**, 959–965 (2011).
34. Zhang, X. *et al.* Cyclic GMP-AMP containing mixed phosphodiester linkages is an endogenous high-affinity ligand for STING. *Mol Cell* **51**, 226–35 (2013).
35. Ablasser, A. *et al.* cGAS produces a 2'-5'-linked cyclic dinucleotide second messenger that activates STING. *Nature* **498**, 380–4 (2013).
36. Diner, E. J. *et al.* The innate immune DNA sensor cGAS produces a noncanonical cyclic dinucleotide that activates human STING. *Cell Rep* **3**, 1355–61 (2013).
37. Gao, P. *et al.* Cyclic [G(2',5')pA(3',5')p] is the metazoan second messenger produced by DNA-activated cyclic GMP-AMP synthase. *Cell* **153**, 1094–107 (2013).
38. Civril, F. *et al.* Structural mechanism of cytosolic DNA sensing by cGAS. *Nature* **498**, 332–7 (2013).
39. Zhang, X. *et al.* The cytosolic DNA sensor cGAS forms an oligomeric complex with DNA and undergoes

- switch-like conformational changes in the activation loop. *Cell Rep* **6**, 421–30 (2014).
40. Li, X. *et al.* Cyclic GMP-AMP synthase is activated by double-stranded DNA-induced oligomerization. *Immunity* **39**, 1019–31 (2013).
  41. Kranzusch, P. J., Lee, A. S., Berger, J. M. & Doudna, J. A. Structure of human cGAS reveals a conserved family of second-messenger enzymes in innate immunity. *Cell Rep* **3**, 1362–8 (2013).
  42. Parvatiyar, K. *et al.* The helicase DDX41 recognizes the bacterial secondary messengers cyclic di-GMP and cyclic di-AMP to activate a type I interferon immune response. *Nat Immunol* **13**, 1155–1161 (2012).
  43. Linder, P. & Fuller-Pace, F. V. Looking back on the birth of DEAD-box RNA helicases. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1829**, 750–755 (2013).
  44. Kadono, M. *et al.* Biological implications of somatic DDX41 p.R525H mutation in acute myeloid leukemia. *Exp Hematol* **44**, 745–754.e4 (2016).
  45. Schütz, P. *et al.* Comparative structural analysis of human DEAD-box RNA helicases. *PLoS One* **5**, e12791 (2010).
  46. Lee, K. G. *et al.* Bruton's tyrosine kinase phosphorylates DDX41 and activates its binding of dsDNA and STING to initiate type 1 interferon response. *Cell Rep.* **10**, 1055–1065 (2015).
  47. Tanaka, Y. & Chen, Z. J. STING specifies IRF3 phosphorylation by TBK1 in the cytosolic DNA signaling pathway. *Sci. Signal.* **5**, ra20 (2012).
  48. Wang, Q. *et al.* Autoinhibition of Bruton's tyrosine kinase (Btk) and activation by soluble inositol hexakisphosphate. *Elife* **4**, 1–31 (2015).
  49. Kabsch, W. *XDS. Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 125–132 (2010).
  50. Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 271–281 (2011).
  51. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. Sect. D Biol. Crystallogr.* **69**, 1204–1214 (2013).
  52. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
  53. Vagin, A. & Teplyakov, A. MOLREP : an automated program for molecular replacement. *J. Appl. Crystallogr.* **30**, 1022–1025 (1997).
  54. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D Biological Crystallography* **66**, 213–221 (2010).
  55. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtand, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 486–501 (2010).
  56. Murshudov, G. N. *et al.* REFMAC 5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 355–367 (2011).
  57. Bryson, K. *et al.* Protein structure prediction servers at University College London. *Nucleic Acids Res.* **33**, 36–38 (2005).
  58. Ward, J. J., McGuffin, L. J., Bryson, K., Buxton, B. F. & Jones, D. T. The DISOPRED server for the prediction of protein disorder. *Bioinformatics.* **20**, 2138–2139 (2004).

59. Omura, H. *et al.* Structural and Functional Analysis of DDX41 : a bispecific immune receptor for DNA and cyclic dinucleotide. *Sci Rep.* **10**, srep34756 (2016).
60. Dürr, H., Körner, C., Müller, M., Hickmann, V. & Hopfner, K. P. X-ray structures of the *Sulfolobus solfataricus* SWI2/SNF2 ATPase core and its complex with DNA. *Cell* **121**, 363–373 (2005).
61. Polprasert, C. *et al.* Inherited and Somatic Defects in DDX41 in Myeloid Neoplasms. *Cancer Cell* **27**, 658–670 (2015).
62. Jiang, Y. *et al.* Structural and functional analyses of human DDX41 DEAD domain. *Protein Cell.* **8**, 72–76 (2017).
63. Rawling, D. C., Fitzgerald, M. E. & Pyle, A. M. Establishing the role of ATP for the function of the RIG-I innate immune sensor. *Elife* **4**, e09391 (2015).
64. Lässig, C. *et al.* ATP hydrolysis by the viral RNA sensor RIG-I prevents unintentional recognition of self-RNA. *Elife* **4**, e10859 (2015).
65. Floor, S. N., Condon, K. J., Sharma, D., Jankowsky, E. & Doudna, J. A. Autoinhibitory Interdomain Interactions and Subfamily-specific Extensions Redefine the Catalytic Core of the Human DEAD-box Protein DDX3. *J. Biol. Chem.* **291**, 2412–2421 (2016).
66. Shimazu, R. *et al.* MD-2, a molecule that confers lipopolysaccharide responsiveness on Toll-like receptor 4. *J Exp Med* **189**, 1777–1782 (1999).
67. Nagai, Y. *et al.* Essential role of MD-2 in LPS responsiveness and TLR4 distribution. *Nat Immunol* **3**, 667–672 (2002).
68. Zhang, Z. *et al.* The E3 ubiquitin ligase TRIM21 negatively regulates the innate immune response to intracellular double-stranded DNA. *Nat Immunol* **14**, 172–178 (2013).
69. Xia, P., Wang, S., Gao, P., Gao, G. & Fan, Z. DNA sensor cGAS-mediated immune recognition. *Protein Cell* **7**, 777–791 (2016).

## Original Paper

1) **Hiroki Omura**, Daisuke Oikawa, Takanori Nakane, Megumi Kato, Ryohei Ishii, Ryuichiro Ishitani,

Fuminori Tokunaga & Osamu Nureki

“Structural and Functional Analysis of DDX41: a bispecific immune receptor for DNA and cyclic dinucleotide”

*Scientific Reports*, 2016; 6; 34756

## Acknowledgements

The author would like to thank Professor Osamu Nureki for his supervision of bachelor, master and Ph.D. courses. In addition, the author also would like to thank Dr. Ryuichiro Ishitani for his supervision of the courses and assistance of writing papers. The author was supported by many members in Nureki laboratory. In particular, the author would like to thank Dr. Ryohei Ishii, Dr. Megumi Kato and Dr. Kazuki Kato for technical advice and experimental support.

During the bachelor's, master's and Ph.D. courses, the author was supported by colleagues in Nureki laboratory. The author would like to thank Mizuki Takemoto, Reiya Taniguchi, Junko Morita, Go Kasuya, Yongchan Lee, Mari Yamada, Masahiro Fukuda and Takashi Yamano for experimental and mental support.

This study was performed in collaboration with Tokunaga laboratory, Osaka City University. The author would like to thank Professor Fuminori Tokunaga and Dr. Daisuke Oikawa for the collaboration in quantitative PCR analysis.

The author also would like to thank Ms. Rieko Yamazaki for secretarial assistance, Arisa Kurabayashi, Sanae Okazaki, Masae Miyazaki and Keiko Ohgomori for technical assistance, the beam-line staff at SLS, Villigen, Switzerland and SPring-8, Hyogo, Japan for technical assistance during data collection.

Finally, the author would like to thank his family for continuous support in his life.